

Rýžování zlata na virtuální obloze z myšlenek vnuknutých Aštarem

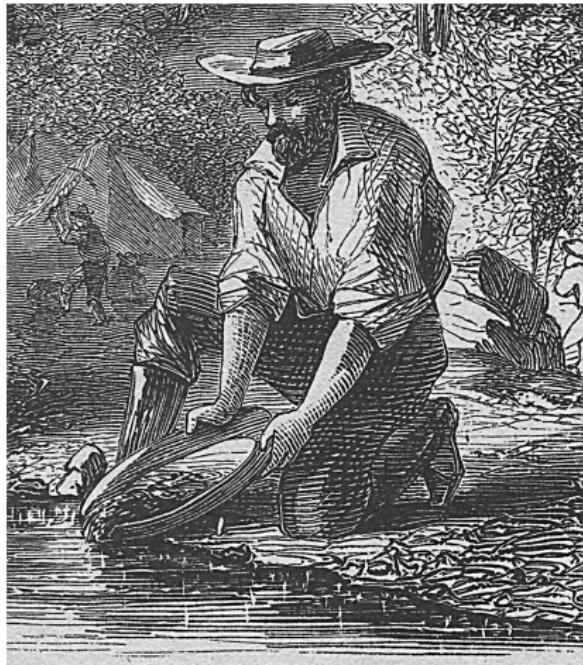
F. Hroch

ÚTFA MU, Brno

14. květen 2010

Rýžování zlata

Gold panning



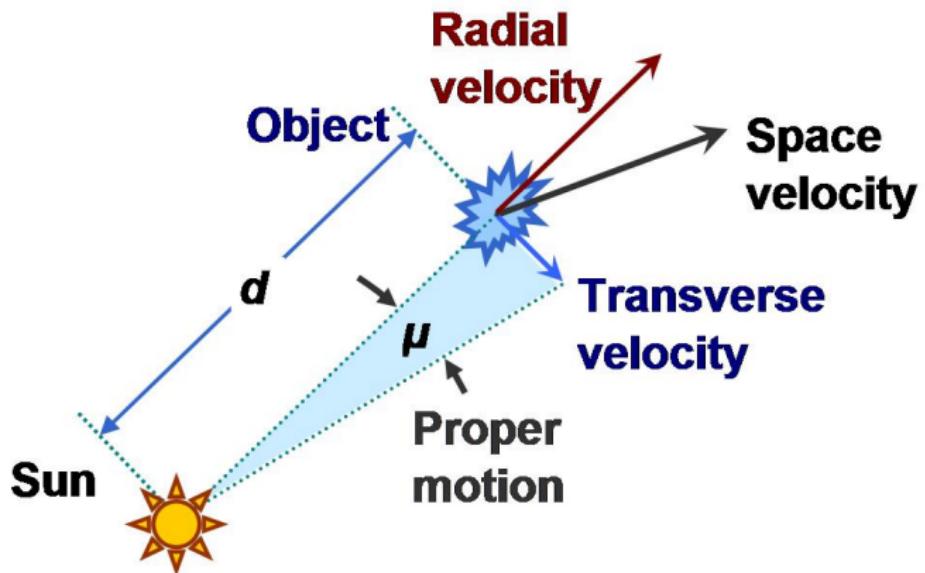
- zlato oddělujeme od vody a písku kroužením v páni, těžší zlato klesá ke dnu
- data a informace jsou rozptýlené v informační zlatonosné řece
- hledání hub: stejný princip, jiná metoda
- nebo vltaviny ...

Vltavíny

Jaro 2010



Hvězda s největším úhlovou rychlostí s největším vlastním pohybem



Přesné polohy z družice Hipparcos

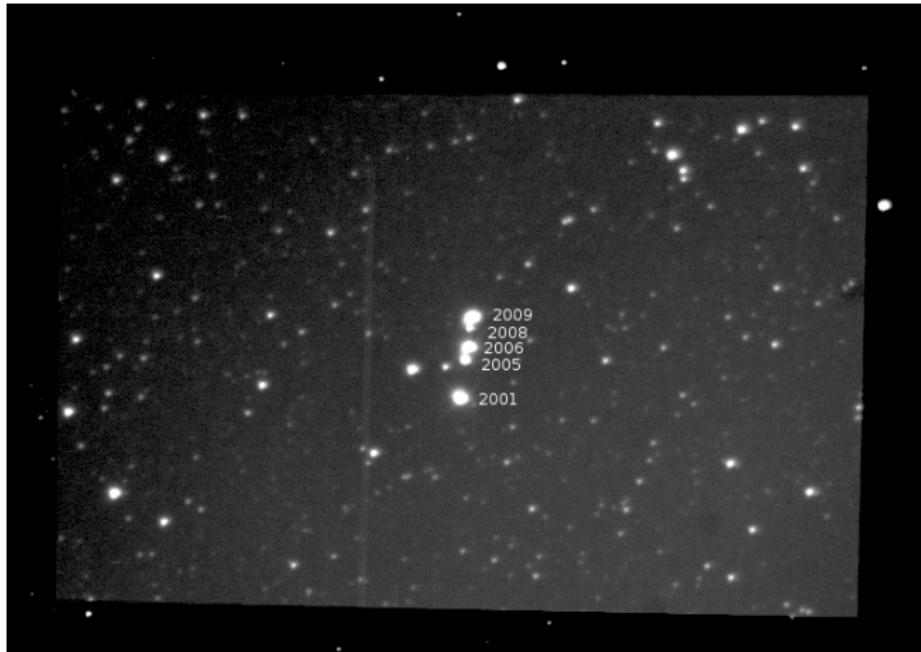
Barnardova hvězda

		mc - /home/hroch.older/cats/hipparcos									
H	638	00 07 48.13 +21 44 44.3	9.86	H 001.95053905 +21.74565092		4.43	-12.74	-31.76	1.09		
H	639	00 07 48.39 -11 07 44.6	8.55	H 001.95164360 -11.12904586		7.06	8.69	-6.18	0.97		
H	640	00 07 50.99 +38 43 18.4	7.66	H 001.96245593 +38.72179052		5.76	-14.56	-23.97	0.60		
H	641	00 07 51.88 +55 34 37.3	7.82	H 001.96614972 +55.57701937		21.83	203.33	8.74	0.55		
H	642	00 07 53.11 -04 51 17.6	9.08	G 001.97130707 -04.85488887		11.55	199.90	-26.97	1.10		
H	643	00 07 54.70 -42 09 13.9 10.10	G 001.97791396 -42.15385091		5.57	-10.57	-44.26	1.10			
H	644	00 07 54.91 -51 15 28.3	8.21	H 001.97877242 -51.25786512		3.50	-22.26	-17.30	0.62		
H	645	H 00 07 59.18 -14 13 36.2	8.85	H 001.99660033 -14.22671354 B		7.79	89.49	-94.73	6.96		
H	647	00 07 59.32 -78 12 39.5	7.79	H 001.99715281 -78.21097486		1.17	25.51	-6.81	0.56		
H	646	00 07 59.34 -18 41 04.9	7.58	H 001.99726573 -18.68470818		3.64	8.65	-7.76	0.91		
H	649	00 08 00.25 -07 32 41.2	7.85	G 002.00105125 -07.54477340		10.66	-28.98	31.53	0.94		
H	648	00 08 00.35 -78 35 55.5	7.18	H 002.00147404 -78.59876143		8.95	-2.21	0.84	0.54		
H	650	00 08 00.76 +29 50 15.0	7.54	H 002.00316528 +29.83750636		3.83	3.29	2.74	0.80		
H	651	00 08 00.79 -39 36 43.5	9.72	H 002.00331222 -39.61207474		6.43	45.91	-17.55	0.80		
H	652	00 08 01.73 +31 23 28.4	7.12	H 002.00721943 +31.39121911		8.19	-28.28	-23.96	0.66		
H	653	00 08 01.89 -30 01 54.1	9.77	H 002.00786726 -30.03169007		6.67	21.70	3.17	1.31		
H	654	00 08 02.05 -45 22 55.7	8.84	H 002.00852236 -45.38213796		4.37	43.52	-15.86	0.91		
H	655	00 08 03.52 -33 31 45.6	5.67	G 002.01465535 -33.52933371		10.02	-42.75	3.15	0.59		
H	656	00 08 04.62 +53 47 47.0	8.15	G 002.01924689 +53.79639636		9.77	73.44	-61.98	0.54		
H	657	00 08 05.02 -29 24 58.9	9.61	H 002.02090072 -29.41637426		3.92	-21.60	-14.55	1.41		
H	658	00 08 05.81 +04 55 38.4	8.19	H 002.02420888 +04.92733420		3.44	-39.04	-36.96	1.14		
H	659	00 08 09.57 +31 22 44.4	7.27	H 002.03986590 +31.37900363		4.90	44.21	-2.18	0.68		
H	660	00 08 10.00 +71 09 34.2	8.51	H 002.04168060 +71.15948850		3.26	3.05	-15.17	0.64		
H	661	00 08 11.72 -59 03 47.1	8.36	H 002.04884213 -59.06308248		2.42	36.25	4.04	0.64		
H	663	00 08 11.94 -38 16 51.8 10.00	G 002.04976404 -38.28106844		4.93	-142.65	-117.19	1.20			
H	662	00 08 11.97 -70 03 21.0	8.74	H 002.04986865 -70.05583226		1.92	7.75	3.04	0.64		
H	664	00 08 12.09 -02 26 51.7	6.18	H 002.05038298 -02.44769895		3.25	7.25	-4.87	0.80		
H	665	00 08 13.16 -43 38 48.1 10.26	G 002.05484372 -43.64669264		7.53	-4.19	-69.58	1.06			
H	666	00 08 13.62 -49 43 18.6 10.16	H 002.05676952 -49.72183630		5.68	6.23	1.99	1.45			
:											

- celkem 118 tisíc hvězd

Barnardova hvězda na MonteBoo

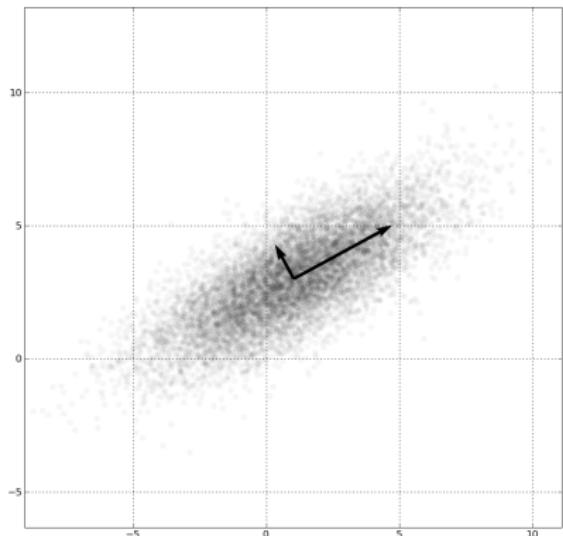
Praktikum z astronomie



- $\mu_\alpha = -0.684$ (-0.79871), $\mu_\delta = 11.660$ (10.33777) ["/year]

Volba správné báze

Neskrývá se něco ve tmě?



- speciální případ: min. – max.
- hledáme přirozenou bázi
- moment setrvačnosti
 $\int_V (\mathbf{r} \cdot \mathbf{r} - \mathbf{r} \otimes \mathbf{r}) \varrho \, dV$
- QR, SVD, ...

Data mining

Rýžování dat



- pátráme po zákonitostech v získaných datech
- data třídíme do skupin (škatulkujeme)
- obvykle matematické nebo statistické nástroje

[10]

Metody pro data mining

katr

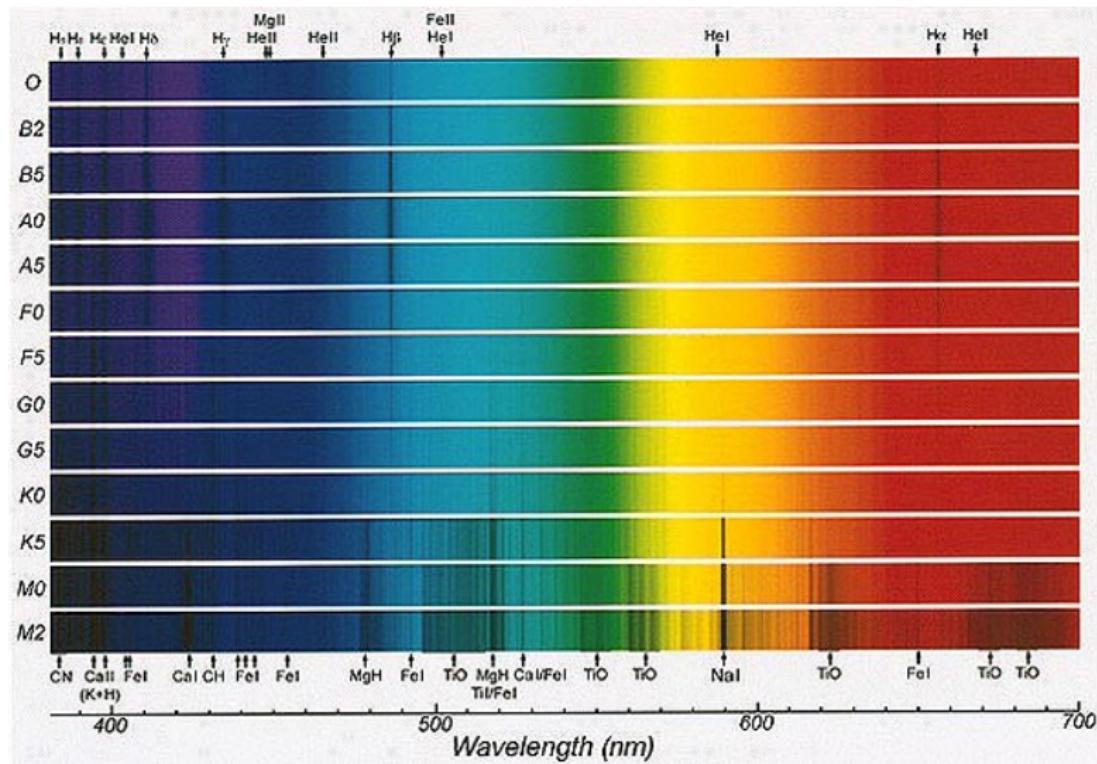
Regresní analýza – použití metody nejmenších čtverců nebo maximální věrohodnosti.

Klasifikace – rozdelení dat do předdefinovaných skupin, volbou vhodné metriky.

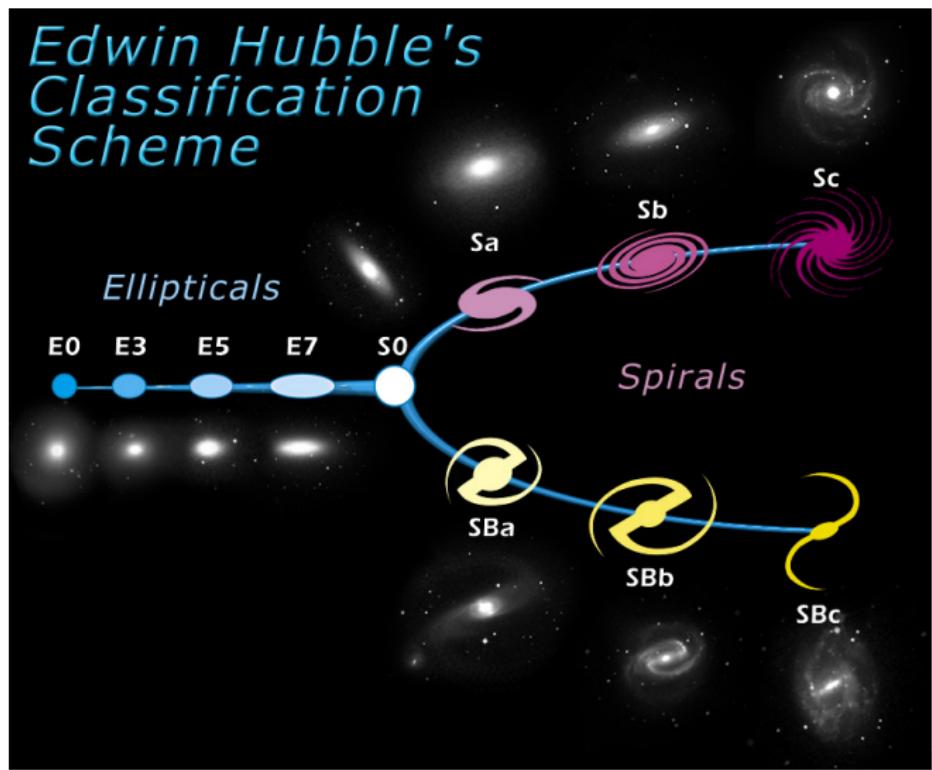
Clustering – Podobná klasifikaci s automatickým určováním skupin a jinými algoritmy.

Analýza nákupního košíku – hledání vzájemných vztahů mezi objekty.

Spektrální klasifikace hvězd



Klasifikace galaxií



Pseudověda?

... nebo hluboké myšlenky?

K čemu je potřeba škatulkování?

- omezené lidské vnímání a myšlení – nepřemýšíme spojité ?
- škatulkování je modelování světa kolem
- zjednodušené vnímání reality?
- Potřebujeme skutečně hvězdné třídy ?
- děláme úředničinu nebo skutečně odhalujeme záhady?

Pseudověda?

- Můj pes má čtyři nohy,
- moje kočka má čtyři nohy,
- moje kočka je tudíž pes...??

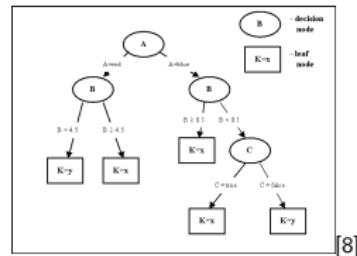
(Sir Arnold z Jistě pane ministře)

Rozhodovací stromy

decision tree

Informační entropie:

$$-\sum_i p_i \log p_i$$

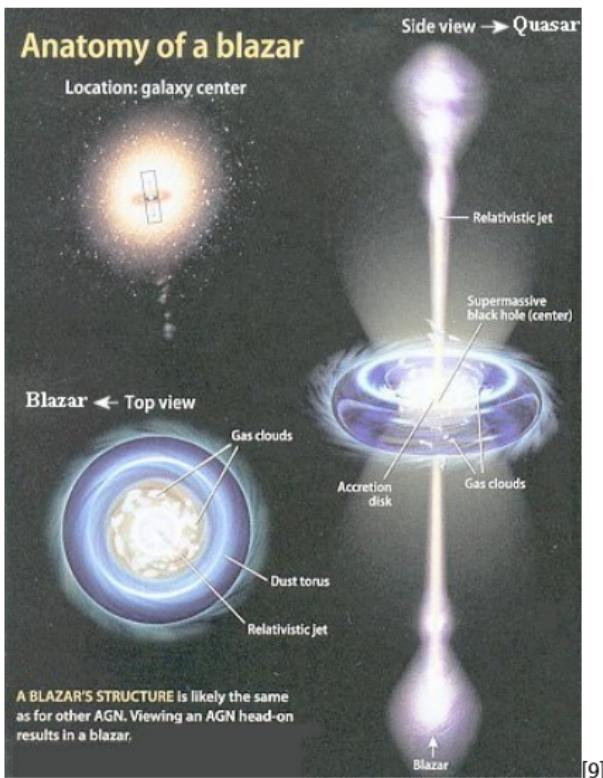


Quinlan's ID3 Algoritmus:

1. Ze všech volných parametrů se spočítá entropie.
2. Vybere se nejmenší
3. Vytvoří se nová větev.
4. Celé se to zopakuje na novou větev.

Hledání blazarů v Sloanově prohlídce oblohy

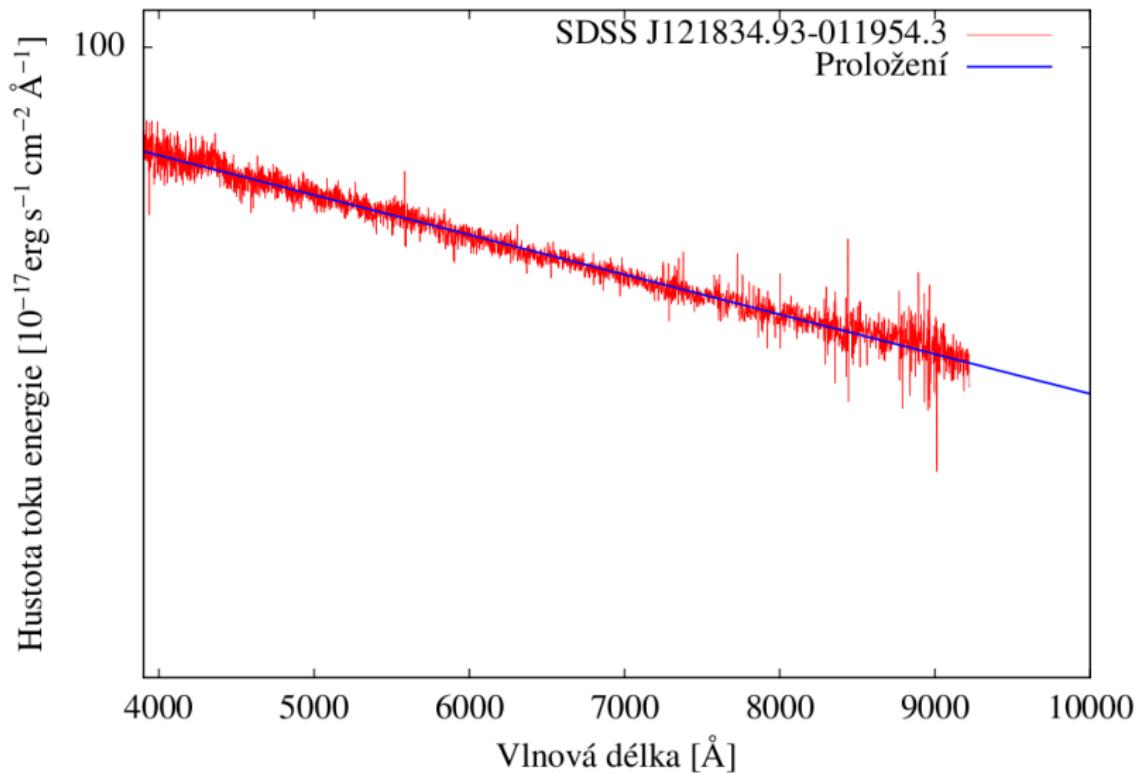
J.Vážný



- bakalářská práce [4]
- Sloan digital sky survey – 2.5 m dalekohled
- synchrotronové záření
- $F(\lambda) \propto \lambda^{-\alpha}$
- robustní fit přímky v $\log F$

Hledání blazarů v Sloanově přehlídce

SDSS J121834.93-011954.3 – Verunka



Hledání zlatých písků

Jak získat hory dat?



- pozorujeme co chceme
- pozorujeme kdy chceme
- potřebujeme vhodný přístroj
- dožijeme se výsledku?

- ihned k dispozici ve velkém množství
- data jsou před-zpracované
- spoléháme na někoho jiného

Jaké data požadujeme?

„Získávání dat pozorováním může být považováno za zastaralé“

Snadná dostupnost — zavrhujeme posílání poštou (diskety, flash memory, děrné štítky), složité podmínky užití.

Předzpracované — obvykle jen osazenstvo příslušné observatoře ví, jak správně zpracovat data bez systematických efektů

Standardní formát — formát je způsob, jak uložit data v počítačových systémech, ukládáme v předem předepsaném a tvaru.

Standardní veličiny — nutná pečlivá kalibrace na konvenčně použitelné veličiny.

Virtuální observatoř

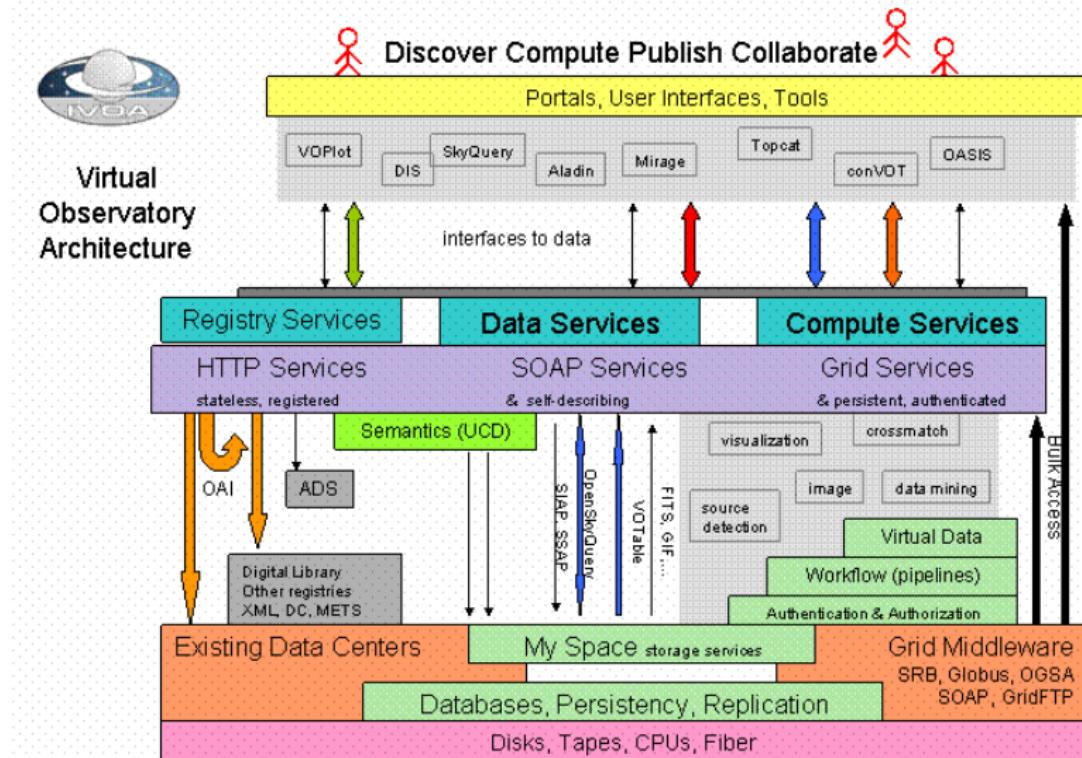
„Dalekohled je jen pomalá databáze“

No image available

- universální přístroj dostupný přes počítač a síť
- poskytuje data ve stejné podobě jako CCD, družice, ...
- k přímému použití astronomem
- neposkytuje nástroje na analýzu

Virtuální observatoř

Schéma



FITS formát

Ukládaní dat (nejen) ve VO

ASTRONOMY & ASTROPHYSICS
SUPPLEMENT SERIES

JUNE 1981, PAGE 363

Astron. Astrophys. Suppl. Ser., 46, (1981) 363-370

FITS : A FLEXIBLE IMAGE TRANSPORT SYSTEM

D. C. WELLS (1), E. W. GREISEN (1) (2) and R. H. HARTEN (4*)

Kit Peak National Observatory Tucson, Arizona, U.S.A.

(1) National Radio Astronomy Observatory Charlottesville, Virginia, U.S.A.

(2) The Netherlands Foundation for Radio Astronomy, Postbus 2, Dwingeloo, the Netherlands

Received March 31, accepted September 4, 1980

Summary.— A format for the interchange of astronomical images and other digital arrays on magnetic tape is described. This format provides a simple but powerful mechanism for the efficient transmission of astronomical images, and also provides a means for the transmission of a virtually unlimited number of auxiliary parameters that may be associated with the image. The parameters are written in a form which is easily interpreted by both humans and computers. The FITS format has been adopted for the transmission of astronomical image data by several large observatories including the Very Large Array, the Westerbork synthesis telescope, the Kitt Peak Observatory and the Anglo-Australian Observatory.

Key words: data analysis

1. Introduction.— With the advent of the VLA in radio astronomy, the increased use of CCD arrays and other digital techniques in optical astronomy, and the development of new observational techniques, the need for a standard way of transmitting astronomical data has become increasingly important. In order to compare images made at different observatories and in order to continue the process of the "image exchange" between the VLA and the WSRT and their home sites or to other observatories, this interchange of data has traditionally been hampered by the fact that each installation has generated its own software system for processing data and its own internal data formats, which differ enormously. Almost every installation has developed at least one unique data format and produced a large quantity of software based on the use of that internal format. Given this

means to transmit arbitrary amounts of text within standard data files in order to encourage the use of "self-documenting" data tapes. Finally, it needs to be general so that it can adapt to new instruments and other inevitable changes, and so that it can be used in image processing applications outside of astronomy.

Over the past several years we have developed and tested a number of formats in an attempt to find a solution to the data interchange problem. These attempts confirmed the validity of the basic idea, but pointed out the dangers of rigidity and complexity in the design of interchange formats. We have reviewed our formats and a variety of other existing formats (for example, "The Flexible Image Transport System", 1977) and concluded that none of them contain adequate simplicity, flexibility or generality. A nonstandard

- Flexible Image Transport System,
- původně rádioastronomický formát,
- ukládá společně data i metadata,
- později rozšířen na obrázky i tabulky.

VOTable

Přenos dat ve VO

- odvozeno z FITS
- obsahuje **metadata** — popisují samotná data
- meta- abstrakce nad něčím
- metadata obrázku: rozměry, souřadnice (pixely – úhly), ...
- metadata tabulky: počet sloupečků, popis jednotlivých sloupečků, jednotky, ...
- cíl typu: program – program
- XML formát
- komunikace prostřednictvím XML-RPC (for hackers only)

Anatomie VOTable

... až na kost

```
<?xml version='1.0'?>
<VOTABLE version="1.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.ivoa.net/xml/VOTable/v1.2"
  xmlns:stc="http://www.ivoa.net/xml/STC/v1.30" >
  <RESOURCE name="myFavouriteGalaxies">
    <TABLE name="results">
      <DESCRIPTION>Velocities and Distance estimations</DESCRIPTION>
      <GROUP ID="J2000" uthype="stc:AstroCoords">
        <PARAM datatype="char" arraysize="" ucd="pos.frame" name="cooframe"
          uthype="stc:AstroCoords.coord_system_id" value="UTC-ICRS-TOPO" />
        <FIELDref ref="col1"/>
        <FIELDref ref="col2"/>
      </GROUP>
      <PARAM name="Telescope" datatype="float" ucd="phys.size;instr.tel"
        unit="m" value="3.6"/>
      <FIELD name="RA" ID="col1" ucd="pos.eq.ra;meta.main" ref="J2000"
        uthype="stc:AstroCoords.Position2D.Value2.C1"
        datatype="float" width="6" precision="2" unit="deg"/>
      <FIELD name="Dec" ID="col2" ucd="pos.eq.dec;meta.main" ref="J2000"
        uthype="stc:AstroCoords.Position2D.Value2.C2"
        datatype="float" width="6" precision="2" unit="deg"/>
      <FIELD name="Name" ID="col3" ucd="meta.id;meta.main"
        datatype="char" arraysize="8"/>
      <FIELD name="Rvel" ID="col4" ucd="spect.dopplerVeloc" datatype="int"
        width="5" unit="km/s"/>
      <FIELD name="e_Rvel" ID="col5" ucd="stat.error;spect.dopplerVeloc"
        datatype="int" width="3" unit="km/s"/>
      <FIELD name="R" ID="col6" ucd="pos.distance;pos.heliocentric"
        datatype="float" width="4" precision="1" unit="Mpc">
        <DESCRIPTION>Distance of Galaxy, assuming H=75km/s/Mpc</DESCRIPTION>
      </FIELD>
      <DATA>
        <TABLEDATA>
          <TR>
            <TD>010.68</TD><TD>+41.27</TD><TD>N 224</TD><TD>-297</TD><TD>5</TD><TD>0.7</TD>
          </TR>
          <TR>
            <TD>287.43</TD><TD>-63.85</TD><TD>N 6744</TD><TD>839</TD><TD>6</TD><TD>10.4</TD>
          </TR>
          <TR>
            <TD>023.48</TD><TD>+30.66</TD><TD>N 598</TD><TD>-182</TD><TD>3</TD><TD>0.7</TD>
          </TR>
        </TABLEDATA>
      </DATA>
    </RESOURCE>
  </VOTABLE>
```

Katalog Hipparcos

očima Virtuální observatoře

Catalogue Selection Page - Chromium

http://vizier.u-strasbg.fr/cgi-bin/vizieR-3

Bookmarks

Query by **Constraints applied on Columns**

Show	Sort	Column	Clear	Constraint	Explain (UCD)
<input type="checkbox"/>	<input type="radio"/>	recno	<input type="button" value="Clear"/>		Record number within the original table (starting from 1) (meta.record) (RECORD)
<input checked="" type="checkbox"/>	<input type="radio"/>	HIP	<input type="button" value="Clear"/>		Identifier (HIP number) (H1) (meta.id:meta.main) (ID_MAIN)
<input type="checkbox"/>	<input type="radio"/>	Proxy	<input type="button" value="Clear"/>	(char)	[HT] Proximity flag (H2) (Note) (meta.note) (REMARKS)
<input checked="" type="checkbox"/>	<input type="radio"/>	RAhms	<input type="button" value="Clear"/>	(char)	Right ascension in h m s, ICRS (J1991.25) (H3) (pos.eq.ra:meta.main) (POS_EQ_RA_MAIN)
<input checked="" type="checkbox"/>	<input type="radio"/>	DEdms	<input type="button" value="Clear"/>	(char)	Declination in deg "", ICRS (J1991.25) (H4) (pos.eq.dec:meta.main) (POS_EQ_DEC_MAIN)
<input checked="" type="checkbox"/>	<input type="radio"/>	Vmag	<input type="button" value="Clear"/>	mag	(n) Magnitude in Johnson V (H5) (phot.mag.em.opt.V) (PHOT_JHN_V)
<input type="checkbox"/>	<input type="radio"/>	VarFlag	<input type="button" value="Clear"/>		(n) [1,3] Coarse variability flag (H6) (Note) (meta.code:src.var) (CODE_VARIAB)
<input type="checkbox"/>	<input type="radio"/>	r_Vmag	<input type="button" value="Clear"/>	(char)	[GHT] Source of magnitude (H7) (Note) (meta.ref) (REFER_CODE)
<input checked="" type="checkbox"/>	<input type="radio"/>	RA(ICRS)	<input type="button" value="Clear"/>	deg	(n) alpha, degrees (ICRS, Epoch=J1991.25) (H8) (Note) (pos.eq.ra:meta.main) (POS_EQ_RA_MAIN)
<input checked="" type="checkbox"/>	<input type="radio"/>	DE(ICRS)	<input type="button" value="Clear"/>	deg	(n) delta, degrees (ICRS, Epoch=J1991.25) (H9) (Note) (pos.eq.dec:meta.main) (POS_EQ_DEC_MAIN)
<input type="checkbox"/>	<input type="radio"/>	AstroRef	<input type="button" value="Clear"/>		[+A-Z] Reference flag for astrometry (H10) (Note) (meta.note) (REMARKS)
<input type="button" value="ALL cols"/> <input type="button" value="Reset All"/> <input type="button" value="Clear"/>					<input type="button" value="Submit Query"/>
<input checked="" type="checkbox"/>	<input type="radio"/>	Pix	<input type="button" value="Clear"/>	mas	(n) indicates a possible blank or NULL column
<input checked="" type="checkbox"/>	<input type="radio"/>	pmRA	<input type="button" value="Clear"/>	mas/yr	(n) Trigonometric parallax (mas) Proper motion in Declination in equatorial coordinates
<input checked="" type="checkbox"/>	<input type="radio"/>	pmDE	<input type="button" value="Clear"/>	mas/yr	(n) Proper motion mu_alpha.cos(delta), ICRS(H12) (for J1991.25 epoch) (pos.pm:pos.eq.ra) (POS_EQ_PMRA)
<input type="checkbox"/>	<input type="radio"/>	e_RAdeg	<input type="button" value="Clear"/>	mas	(n) Proper motion mu_delta, ICRS (H13) (for J1991.25 epoch) (pos.pm:pos.eq.dec) (POS_EQ_PMDEC)
<input type="checkbox"/>	<input type="radio"/>	e_DEdeg	<input type="button" value="Clear"/>	mas	(n) Standard error in RA*cos(DEdeg) (H14) (at epoch J1991.25; for different epochs, the actual mean error must take into account the proper motion uncertainties) (stat.error) (ERROR)
<input checked="" type="checkbox"/>	<input type="radio"/>	e_Pix	<input type="button" value="Clear"/>	mas	(n) Standard error in DE (H15) (at epoch J1991.25; for different epochs, the actual mean error must take into account the proper motion uncertainties) (stat.error) (ERROR)
<input type="checkbox"/>	<input type="radio"/>	e_pmRA	<input type="button" value="Clear"/>	mas/yr	(n) Standard error in pmRA (H17) (stat.error) (ERROR)
<input type="checkbox"/>	<input type="radio"/>	e_pmDE	<input type="button" value="Clear"/>	mas/yr	(n) Standard error in pmDE (H18) (stat.error) (ERROR)
<input type="checkbox"/>	<input type="radio"/>	DE:RA	<input type="button" value="Clear"/>		(n) [-1/1] Correlation, DE/RA*cos(delta) (H19) (stat.correlation) (STAT_CORRELATION)
<input type="checkbox"/>	<input type="radio"/>	Pix:RA	<input type="button" value="Clear"/>		(n) [-1/1] Correlation, Pix/RA*cos(delta) (H20) (stat.correlation) (STAT_CORRELATION)
<input type="checkbox"/>	<input type="radio"/>	Pix:DE	<input type="button" value="Clear"/>		(n) [-1/1] Correlation, Pix/DE (H21) (stat.correlation) (STAT_CORRELATION)
<input type="checkbox"/>	<input type="radio"/>	pmRA:RA	<input type="button" value="Clear"/>		(n) [-1/1] Correlation, pmRA/RA*cos(delta) (H22) (stat.correlation) (STAT_CORRELATION)
<input type="checkbox"/>	<input type="radio"/>	pmRA:DE	<input type="button" value="Clear"/>		(n) [-1/1] Correlation, pmRA/DE (H23) (stat.correlation) (STAT_CORRELATION)

Jednotná data

Astrometrie, spektroskopie

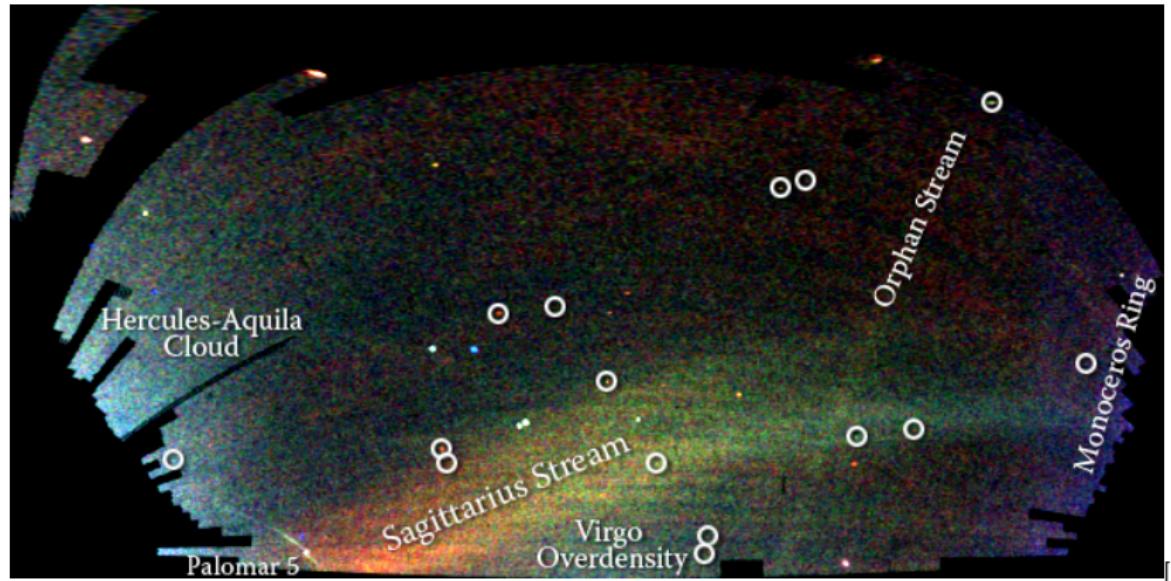
- existuje jednotný standard — pixely, α, δ, λ

Fotometrie

- ve 21. století nemáme sjednocené fotometrické veličiny
- radio — W/m^2 , Jansky
- optika, IR — magnitudy: $m = m_0 - 2.5 \log_{10} I/I_0$
- RTG, γ — keV, MeV

Masivní práce s Virtuální observatoří

hvězdné proudy v naší galaxii (Sloanova prohlídka)



Klasika a rýžování

Když dva dělají totéž

Klasika

- Výběr sympatického objektu
- noci strávené u dalekohledu
- zpracování naměřených dat
- model
- závěry

Rýžování

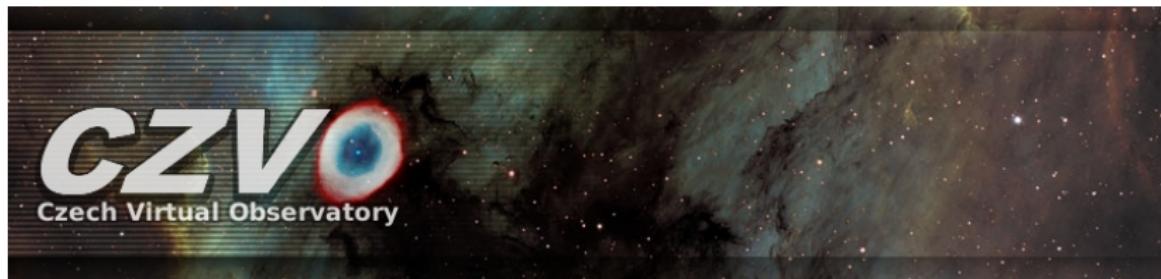
- pátráme po něčem abstraktním
- noci strávené u počítače
- model
- závěry

Použití Virtuální observatoře

- další druh přístroje
 - snadná vizualizace dat
 - snadné spojení několika informačních zdrojů
 - snadné předávání dat mezi programy
 - homogenizace dat
-
- studium vlastních pohybů (hvězdokupy, asociace, . . .)
 - sestavování multispektrálních obrázků a spekter
 - objevování nových objektů (hnědí a bílí trpaslíci, blazary, . . .)
 - studium velkoškálových struktur ve vesmíru

Česká virtuální observatoř

Czech Virtual Observatory



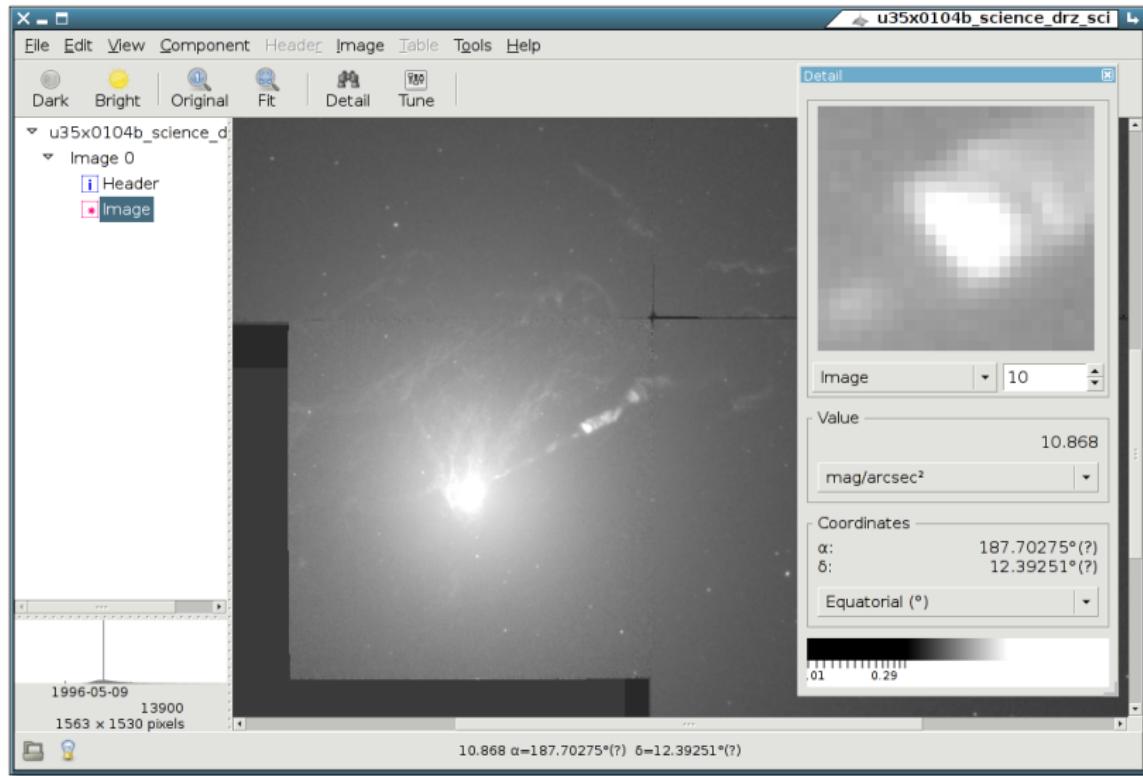
- <http://stelweb.asu.cas.cz/czvo/>
- duchovní otec: Petr Škoda
- Ondřejov: centrum, infrastruktura, spektra
- Brno: metody, data mining, software

Virtuální archív Monteboo

Night	List of targets	Info	References
20100107	ZZ Cas, zzcas, dark, stper, zz cas		
20100112	xyuma, dark		
20100116	nsvs0745, dark		
20100120	dark		
20100125	dark		
20100126	avcmi, mars, uxuma, dark, flat		
20100129	uuand, dark		
20100131	avcmi, dark, adand, ttaur, m1, ft_ori, et_boo		
20100201	lv_cas		
20100202	adand, v471tau, v471tau2, lmaur, dark		
20100203	fg_gem, NSVS1031772, dark		
20100207	dark		
20100210	dark		
20100221	V471tau, V582mon, avcmi, dark, m61, saturn, m64, cvdra, cvdra2, uxuma, m63, flat		
20100222	ft_gem, dark		
20100225	UU Lyn, dark, Moon, avcmi, avcmi2		
20100227	AV CMi, dark, LP UMa		
20100301	ag_leo, dark		
20100302	gu_boo, lv_uma, dark		
20100303	ao_mon, AG Leo, dark		

přes 150 tisíc snímků, 5.6 mil. sekund (65 dní) ... za 5 let

Podpora VO pro Munipack

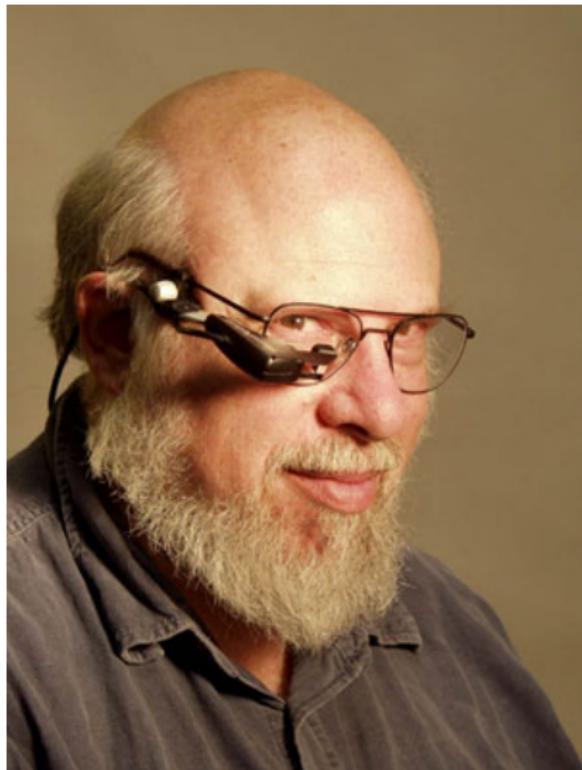


Zdroje informací

-  <http://www.physics.muni.cz/~hroch/ryzovani.pdf>
-  <http://stelweb.asu.cas.cz/czvo/>
-  <http://www.ivoa.net>
-  <http://www.euro-vo.org/pub/>
-  http://is.muni.cz/th/211665/prif_b/bak.pdf
-  <http://www.abclinuxu.cz/clanky/rozhovory/jef-raskin-programmers-at-work>

Jef Raskin

1943 — 2005



- matematik, hudebník, ...
- otec Apple Macintosh
- autor The Humane Interface
- počítače jako spotřebiče
- metafory
- myšlenkové modely
- rýžování ...
- ... a fyzika ?

Reference

-  http://cs.wikipedia.org/wiki/Rýžování_zlata
-  http://en.wikipedia.org/wiki/Data_mining
-  http://en.wikipedia.org/wiki/C4.5_algorithm
-  <http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch-20040615.html>
-  http://en.wikipedia.org/wiki/Proper_motion
-  odin.physastro.mnnsu.edu
-  <http://en.wikipedia.org/wiki/File:HubbleTuningFork.jpg>
-  http://dms.irb.hr/tutorial/tut_dtrees.php
-  <http://universe-review.ca/F05-galaxy.htm>
-  <http://thomaslarock.com/2009/06/sql-2008-data-mining/>
-  http://www.sdss.org/includes/sideimages/fos_dr6_marked.html